

CONCEPTUALISATION DE L'ANALYSE STATISTIQUE IMPLICATIVE

Jean-Claude Oriol* & Jean-Claude Régnier**

*CERRAL IUT LUMIERE & EA 3727 Université Lyon2

**Enseignant-Chercheur EA3727 Université Lyon2

Université Lumière Lyon II 16 Quai Claude Bernard F69007 Lyon

Résumé : ce texte se propose de décrire une activité visant à permettre à des étudiants de deuxième année du Diplôme Universitaire de Technologie STID (Statistique et Traitement Informatique des Données) d'utiliser l'analyse statistique implicative. Nous donnerons en particulier les obstacles rencontrés concernant la conceptualisation des méthodes d'analyse implicative et les outils mis en place afin de les surmonter.

Abstract : this text proposes to describe an activity aiming at making it possible students of 2nd year of the University Diploma of Technology STID (Statistical and Data Processing) to use the implicative statistical analysis. We will give in particular the obstacles met concerning the conceptualization of the methods of implicative analysis and the tools installed in order to surmount them

Mots clés : statistique, conceptualisation, apprentissage, didactique de la statistique, analyse implicative.

1. L'analyse statistique implicative : une très brève introduction

1.1. Premiers pas

L'analyse statistique implicative vise à donner du sens à des énoncés du type $a \Rightarrow b$ lorsque l'implication n'est pas "totale". Nous faisons référence ici à l'analyse définie par Régis Gras (1996) sans méconnaître d'autres approches de la classification, des arbres hiérarchiques et du partitionnement que l'on trouvera par exemple chez Nakache et Confais (2005) qui ne manquent ni d'intérêt ni de pertinence mais qui ne sont pas notre propos actuel.

Dans l'analyse statistique implicative Régis Gras (1996) définit plusieurs notions sous la forme d'indices ou fonctions, et les deux premières notions abordées sont celles d'indice d'implication et d'intensité d'implication.

La modélisation de la situation est de considérer dans un premier temps des variables binaires a et b qui sont réalisées ou non pour chaque individu d'une population. Soit n la taille de la population, n_a le nombre d'individus où a est réalisé, $n_{\bar{a}}$ le nombre où a n'est pas réalisé, etc. Avec des notations similaires on aura :

$n_{a \cap \bar{b}}$ = nombre d'individus où a est vrai et b est faux.

1.2. La notion d'indice d'implication

Dans ces conditions l'indice d'implication est défini par l'expression suivante :

$$q(a, \bar{b}) = \frac{n_{a \cap \bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$$

On aura tout de suite remarqué l'utilisation du « facteur » d'indépendance entre a et $\text{non}(b)$.

1.3. L'intensité d'implication

La variable aléatoire correspondante $Q(a, \bar{b})$ peut s'approximer par une loi normale dans les cas où l'on peut justifier cette approximation c'est-à-dire si $n_a * n_{\bar{b}} > 3n$, et $n_a < n_b$ et $n_b \neq n$. Dans ce cas on peut définir une espèce de valeur de l'admissibilité de l'implication $a \Rightarrow b$ par l'intensité d'implication qui est une fonction définie de la façon suivante :

$$\varphi(a, \bar{b}) = 1 - \Pr [Q(a, \bar{b}) \leq q(a, \bar{b})] = \frac{1}{\sqrt{2\Pi}} \int_{q(a, \bar{b})}^{\infty} e^{\frac{-t^2}{2}} dt$$

Nous pourrions à loisir développer les diverses notions rencontrées dans l'analyse statistique implicative et leurs propriétés mais nous allons figer ici notre introduction afin d'exposer le contexte de cette étude.

2. Le contexte de l'IUT Lumière et de la statistique dans le département STID

2.1. L'alternance à l'IUT Lumière

Depuis sa création en 1992 l'IUT Lumière est le seul IUT dans lequel la formation des étudiants est complètement réalisée en alternance. Cette formation est ainsi construite : en première année les étudiants suivent un stage de sept semaines dans une entreprise pendant la période avril-mai, la deuxième année est réellement en alternance sur la base d'un contrat d'apprentissage. La modalité en place dans le département STID est de 15 jours à l'IUT suivis

de 15 jours en entreprise. L'année universitaire s'achève par la tenue des jurys et la délivrance des diplômes au mois de septembre. Cette forme d'organisation des études est usuellement désignée par «système 1+1».

2.2. Les projets tuteurés

Dans l'organisation des enseignements de statistique une cinquantaine d'heures en deuxième année est consacrée à un projet appelé « Projet d'études statistiques ».

Les étudiants par groupes de cinq ou six doivent conduire une étude encadrée par un enseignant pour un commanditaire extérieur.

Deux années consécutives des étudiants du département ont mené une étude sur les représentations pour l'équipe de recherche en sciences de l'éducation ADATIC, avec comme consigne de « mettre en actes » l'implication statistique implicative et en utilisant le logiciel CHIC construit spécialement pour de telles études.

3. Le cadre théorique de l'apprentissage

Le point de vue de la formation en alternance nous renvoie à la question fondamentale : Comment apprenons-nous ?

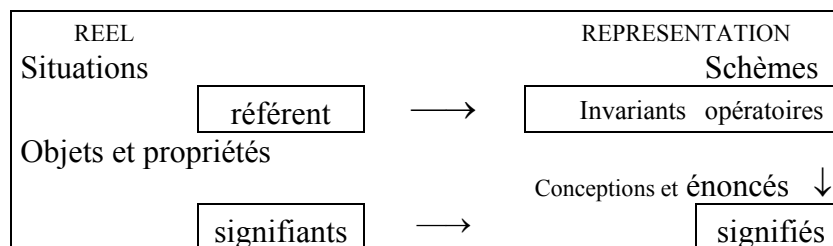
La théorie des champs conceptuels de Gérard Vergnaud (1991) p.135 répond de manière satisfaisante pour ce qui touche aux apprentissages en statistique : « *La théorie des champs conceptuels est une théorie cognitive, qui vise à fournir un cadre cohérent et quelques principes de base pour l'étude du développement et l'apprentissage des compétences complexes, notamment celles qui relèvent des sciences et des techniques.(...) [et] pour la compréhension des filiations et des ruptures entre connaissances* ».

La conceptualisation, élément central de l'apprentissage, prend appui sur le concept conçu comme « un triplet de trois ensembles référence, signifié et signifiant :

- *l'ensemble des situations qui donnent du sens au concept (la référence),*
- *l'ensemble des invariants sur lesquels repose l'opérationnalité des schèmes (le signifié)*
- *l'ensemble des formes langagières et non langagières qui permettent de représenter symboliquement le concept, ses propriétés, les procédures de traitement (le signifiant) »*

Suivons Gérard Vergnaud (1994), p. 180, qui pour affiner « *progressivement la définition d'un schème,...[dit]... d'abord que c'est une totalité dynamique fonctionnelle, c'est-à-dire quelque chose qui fonctionne comme une unité ; en second lieu que c'est une organisation invariante de la conduite pour une classe de situations données (l'algorithme est un cas particulier du schème) ; et en troisième lieu qu'un schème est composé de quatre catégories d'éléments : 1)des buts, intentions et anticipations ; 2)des règles d'action; 3)des invariants opératoires; 4)des possibilités d'inférences en situation.* »

Revenant sur ce qu'il appelle lui-même le « triangle des linguistes », Vergnaud (1994), p. 189-190 propose le modèle suivant :



Ainsi la notion de champ conceptuel en tant qu'ensemble des situations renvoyant à l'idée de procédure, permet de situer la simulation en statistique dans sa finalité qui est de fournir à l'apprenant un éclairage sur le signifié (invariants opératoires)

Par ailleurs l'approche développée par Jean-Claude Régnier (1988) intégrant l'apprentissage fondé sur le tâtonnement expérimental de l'apprenant offre une perspective pour analyse didactique de la simulation et relation à la résolution de problème.

4. La simulation outil d'appropriation de concepts statistiques

Nous avons débattu de l'utilité et de l'utilisation de la simulation dans diverses situations rencontrées dans l'enseignement de la statistique entre autres concernant les intervalles de confiance (Oriol et Régnier 2004) et le coefficient de corrélation (Oriol et Régnier 2006).

C'est sans doute un des points spécifiques de l'enseignement de la statistique obligé de développer un enseignement s'appuyant sur les mathématiques mais hétérodoxe par rapport aux outils traditionnels construits dans l'exclusion entre le vrai et le faux. Notons d'ailleurs que dans notre contexte de l'analyse statistique implicative la pensée développée est « doublement » hétérodoxe : d'une part comme toute pensée statistique et d'autre part comme s'intéressant à des énoncés $a \Rightarrow b$ « partiellement » vrais.

5. La situation proprement dite

5.1 Le choix d'Excel

Notre pratique pédagogique vise à intégrer le plus tôt possible l'outil informatique comme instrument canonique d'une pratique de la statistique. Et en matière d'outil informatique, le tableur est privilégié dans un premier temps. Les cours « spécialisés » à l'aide de logiciels tels que SAS, SPAD complètent la gamme des compétences de nos étudiants.

5.2 La construction de la simulation

La séance didactique prend appui sur une bonne maîtrise du logiciel Excel par les étudiants et fait suite à une séance de deux heures sur l'approche de l'analyse statistique implicative.

Les étudiants doivent construire une feuille Excel tirant 100 fois au hasard les valeurs binaires de a et b, évaluer l'indépendance des variables a et b, calculer pour ces 100 valeurs l'indice d'implication et l'intensité d'implication entre a et b.

Et recommencer...

Voici les 10 premières lignes du tableau

	a	b						verif
			16	31	39	14	100	100
1	1	1	0	0	0	1	1	
2	1	0	0	0	1	0	1	
3	1	0	0	0	1	0	1	
4	0	1	0	1	0	0	1	
5	1	1	0	0	0	1	1	
6	1	1	0	0	0	1	1	
7	1	0	0	0	1	0	1	
8	1	0	0	0	1	0	1	
9	0	0	1	0	0	0	1	
10	0	0	1	0	0	0	1	

Un des tableaux de contingence obtenus

a --				
>		0	1	
b v	0	16	39	55
	1	31	14	45
		47	53	100

Le tableau de contingence et le test du Khi deux

Tableau d'indépendance

25,85 29,15
21,15 23,85

Calcul du Xhi deux

3,75329 3,3284
4,58735 4,068

15,7370574

Le calcul de l'indice d'implication et de l'intensité d'implication

Indice d'implication

0,2776

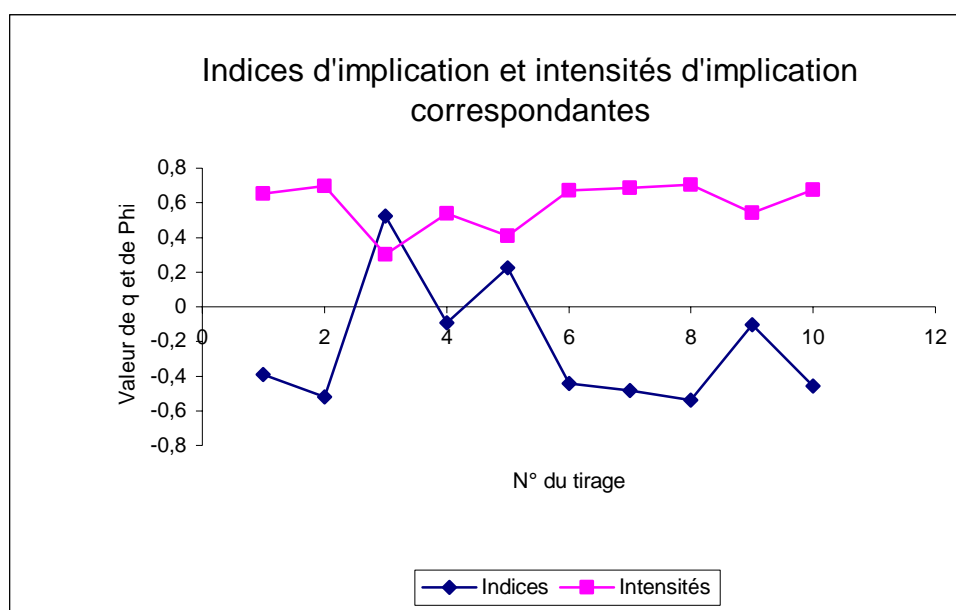
Intensité d'implication

0,39066

Voici une série de résultats obtenus concernant des indice d'implication et les valeurs d'intensités correspondantes

1	-0,390457	0,65190056
2	-0,519615	0,69833413
3	0,5223864	0,30070065
4	-0,092884	0,53700223
5	0,226187	0,41052805
6	-0,442627	0,67098208
7	-0,482015	0,68510217
8	-0,539036	0,70506905
9	-0,104787	0,54172779
10	-0,457265	0,67625955

Et une représentation graphique :



Les remarques évidentes à la lecture des dizaines de graphiques obtenus, par exemple si l'indice q est inférieur à 0 alors l'intensité est supérieure à 0,5 vont permettre de passer aux démonstrations de ces propriétés puis à poursuivre et à aborder les concepts suivants du champ de l'analyse statistique implicite.

6. Conclusion

Il nous apparaît que la construction par les apprenants de l'outil de simulation est garante d'une meilleure appropriation des concepts statistiques. Nous pouvons également signaler que parmi les problèmes en suspens reste le contrôle par les étudiants du générateur de nombres aléatoires d'Excel. Nous développerons à l'oral les résultats attendus, espérés et observés de la variation de l'énoncé de l'activité proposée, puis comment cette activité de simulation trouve une contextualisation dans la pratique d'enquêtes réalisées par les étudiants dans la suite de ces travaux tuteurés.

7. Bibliographie

1. Albert, J. H. (1993), "Teaching Bayesian Statistics Using Sampling Methods and MINITAB," *The American Statistician*, 47, 182-191.
2. Antoine F., Grootaers D., Tilman F., (1988), Manuel de la formation en alternance, Lyon, Chronique sociale : Bruxelles, Vie ouvrière.
3. Bachelard G., (1938), La formation de l'esprit scientifique, Vrin.
4. Batareno C. et Godino J., (2001), Análisis de Datos y su Didáctica, Universidad de Granada
5. Brousseau G. (1998), *Théorie des situations didactiques*, La Pensée Sauvage Editions, Grenoble.
6. Gras R. (1996), *L'implication statistique*, La Pensée Sauvage Editions, Grenoble.
7. Hesterberg, T. C. (1998), "Simulation and Bootstrapping for Teaching Statistics," *American Statistical Association Proceedings of the Section on Statistical Education*, Alexandria, VA: American Statistical Association, 44-52.
8. Kennedy, K., Olinsky, A., and Schumacher, P. (1990), "Using Simulation as an Integrated Teaching Tool in the Mathematics Classroom," *Education*, 111, 275-296.
9. Le Nir M., Oriol JC, (2000), L'enquête : élément central de l'enseignement de statistique en première année GLT in CNR'IUT 2000 Recherche et innovation, Presses Universitaires d'Orléans, pages 157 à 169.

10. Malglaive G., Weber A., (1982), « Théorie et pratique : approche critique de l'alternance en pédagogie », *Revue française de pédagogie*, n°61, pp. 17-27.
11. Marasinghe, M. G., Meeker, W. Q., Cook, D., and Shin, T. (1996), "Using Graphics and Simulation to Teach Statistical Concepts," *The American Statistician*, 50, 342-351.
12. Mills, J. D., (2002), Using Computer Simulation Methods to Teach Statistics: A Review of the Literature, *Journal of Statistics Education* Volume 10, Number 1
13. Nakache J.-P. et Confais J. (2005), *Approche pragmatique de la classification*, Editions Technip, Paris.
14. Oriol J-C, (1999), "*L'évaluation en alternance*", in Actes du colloque sur l'alternance, IUT Lumière, Lyon.
15. Oriol J-C, (2002), "*Réaliser une enquête par questionnaires : un outil didactique pour la statistique inférentielle à l'Université*", in 3^{ème} Rencontre Europe-Amérique-Latine sur la formation technologique et professionnelle, La Havane, Cuba.
16. Oriol JC, Régnier JC (2003), *Fonctionnement didactique de la simulation en statistique : Exemple de l'enseignement du concept de corrélation*, Colloque EMF 2003, Tozeur
17. Oriol JC, Régnier JC (2003), *La loi de Benford et la pratique d'enquêtes : un exemple de Situations Didactiques en Statistique*, Montpellier, SFSD 2004
18. Pedhazur, E. J. (1997), *Multiple Regression in Behavioral Research: Explanation and Prediction* (3rd ed.), Orlando: Harcourt Brace College.
19. Raynal F., Reunier A., (1997), *Pédagogie : dictionnaire des concepts clés*, ESF Editeur, Paris, 408 p.
20. Régnier, JC, (1983) Étude didactique d'un test autocorrectif en trigonométrie, Strasbourg : IREM-Université L Pasteur, Tome 1 : 307 pp, Tome 2 : 171 p Thèse de doctorat de 3ème cycle en didactique des mathématiques
21. Régnier, J-C, (1988) Étude didactique d'une méthode d'apprentissage fondé sur le tâtonnement expérimental de l'apprenant, *Annales de Didactique et de Sciences Cognitives*, séminaire de Didactique des Mathématiques de Strasbourg, pp 255-279
22. Régnier, JC, (1988) Étude didactique d'une méthode d'apprentissage fondé sur le tâtonnement expérimental de l'apprenant, *Annales de Didactique et de Sciences Cognitives*, séminaire de Didactique des Mathématiques de Strasbourg , pp 255-279
23. Régnier, JC, (1994), *Statistique documents auto-correctifs et auto-évaluatifs d'aide à l'apprentissage*, ISPEF, Université Lyon2.
24. Régnier, JC, (2000a) L'alternance: des évidences (en) aux questions. Forum, revue de la recherche en travail social, n°93, pp 38-47 ISSN 0988.6486
25. Régnier, JC, (2000b) Auto-évaluation et autocorrection dans l'enseignement des mathématiques et de la statistique. Note de synthèse pour l'obtention de l'Habilitation à Diriger des Recherches. Université Marc Bloch (13 décembre 2000) 240 p.
26. Revuz A., (1980), *Est-il impossible d'enseigner les mathématiques ?*, Puf.
27. Ronan C., (1988), *Histoire mondiale des sciences*, Seuil, 704 p.
28. Vergnaud, G., (1991) La théorie des champs conceptuels, *Recherches en Didactique des mathématiques*, 10/2.3, pp. 133-169.
29. Vergnaud, G., (1994) Le rôle de l'enseignant à la lumière des concepts de schème et de champ conceptuel, in M. Artigues, R. Gras, C. Laborde, P. Tavignot (EDS),
30. Vingt ans de didactiques des mathématiques en France, Grenoble, La Pensée Sauvage Editeurs, pp. 177-191.